# Development of High-Resolution Acoustic Camera based Real-Time Object Recognition System by using Autonomous Underwater Vehicles.

Son-Cheol Yu, Tae-Won Kim, Akira Asada*, Scott Weatherwax, Ben Collins and Junku Yuh**
University of Hawaii
*University of Tokyo
**National Science Foundation

***Abstract*-This paper addresses acoustic camera DIDSON based object recognition method for AUVs(Autonomous Underwater Vehicles). The acoustic camera's characteristics and display method based on various experiments and the acoustic camera model and efficient and reliable image recognition method for AUVs are proposed. As examples, the cubic and cylindrical objects modeled to simulate the image and their recognition test was carried out in experimental tank.**

## I. INTRODUCTION

AUVs have been successfully automated for long range data collection missions and have been proposed for use in risky and complicated missions, that would otherwise be hazardous to humans, such as safety inspections [1][2][3]. However, due to limited object recognition capability, underwater applications for AUV's are still limited. As a result, development of underwater object recognition is key to the emergence of AUVs as a viable resource in solving problems underwater.

Image sonar and optical camera systems have been widely used for object recognition. However, several major disadvantages have hindered their practical implementation. Imaging sonar's range is relatively long and reliable but higher resolution than what has been traditionally available is required for use in object recognition [4]. Alternatively, optical camera systems have been shown to have the required resolution, but the limited visibility encountered in underwater environments restricts working range and reliability [5].

The high-resolution acoustic cameras such as DIDSON [6][7][8][9] can be an alternative. It has the required range and reliability needed for optical recognition combined with the high resolution seen in traditional optical vision systems. However, the acoustic camera has unique characteristics that hinder its use in autonomous object recognition tasks.

In the section 2, the acoustic camera's characteristics and the differences with respect to traditional optical cameras are studied. In the section 3, a proposed camera model recognition method is addressed. The section 4 and 5 will give simulation and experimental results which applied the model proposed in section 3.

## II. CHARACTERISTICS OF THE ACOUSTIC CAMERA

The acoustic camera system has several defining features that make image recognition more challenging when compared to a traditional optical camera system. Figs 1 to 12 show images taken using the DIDSON acoustic camera. The DIDSON camera (1.8Mhz) captured various objects on a plain surface at about 1m distance. There are large differences between the produced acoustic images when compared to traditional optical images of the same objects. These differences are due in large part to the acoustic nature of DIDSON and are highlighted below:

1*) Display Method:* Its display method is different from the optical camera. As shown in Fig.13, the reflected light from the object and background maps to the CCD using the corresponding line which connects the reflection spot and focus point. In this situation there is one to one mapping: at no point do reflections from the object and from the background behind the object overlap on the CCD. This is not the case with the DIDSON acoustic camera.

As shown in Fig.14, the acoustic camera emits acoustic beams (from point *DP*) and returns two sets of data, the intensity of the return from a point, *I,* and the distance from the camera, point *DP*, to the reflection point, *D*.

The difference in images produced by an acoustic and an optical camera occurs when the acoustic return, a function of *I* and *D*, gets mapped into an image [10]. Fig. 14 models the acoustic image produced by simple cylinder, with the upper right image seen in Fig. 14, displaying the full image produced by the acoustic camera. In this image, the top of the cylinder, *C0*, maps to the bottom of the acoustic image and the side of the cylinder, *C1,* maps to the middle of it. This produces an "upside down" version of the same object as viewed with a traditional optical camera.

This geometry is due to the shorter distance *D* between points *DP* and *C0* compared to the distance *D* between points *DP* and *C1*. If the two beams have the same distances *D*, their intensities *I* overlap on the same point on the acoustic image. An example of this effect is visible in Fig.14. Due to the relative distances, the intensities *I* of *B1*, part of *C1* and *S1* overlapped to the same area. As the intensities overlap the area of overlap becomes darker. The result of this overlap is partly deformed or lost objects in the image. As the acoustic camera position changes the image, depending on the object height and the camera position, changes as well. Of notable

1-4244-0115-1/06/$20.00 ©2006 IEEE

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**01 SEP 2006** | 2. REPORT TYPE<br>**N/A** | 3. DATES COVERED<br>**-** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**Development of High-Resolution Acoustic Camera based Real-Time Object Recognition System by using Autonomous Underwater Vehicles** | 5a. CONTRACT NUMBER<br>5b. GRANT NUMBER<br>5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER<br>5e. TASK NUMBER<br>5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**University of Hawaii** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release, distribution unlimited**

13. SUPPLEMENTARY NOTES
**See also ADM002006. Proceedings of the MTS/IEEE OCEANS 2006 Boston Conference and Exhibition Held in Boston, Massachusetts on September 15-21, 2006, The original document contains color images.**

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **UU** | **6** | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

Fig.1 Brick
(9.1x19.3x5.8(H)cm)

Fig.2 Concrete Cube
(17.8x17.8x10.2(H)cm)

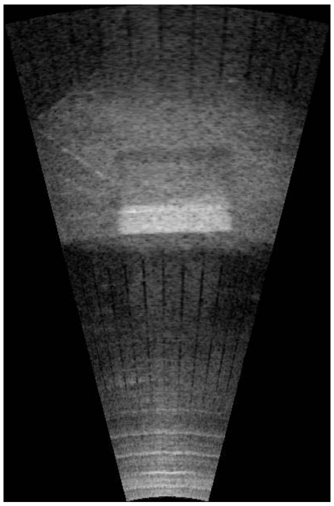

Fig.3 Brick Image

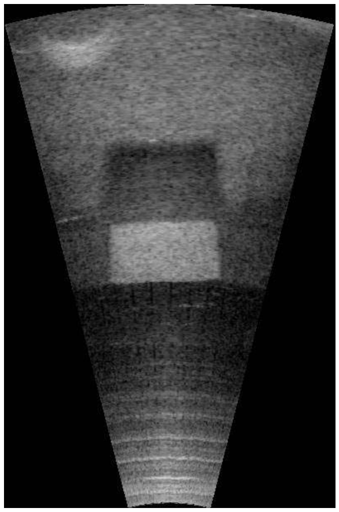

Fig.4 Concrete Cube Image

Fig.5 Concrete Cylinder
(15(D) x 30.6(H)cm)

Fig.6 Metal Cylinder
(12.7(D) x 34.1(H)cm)

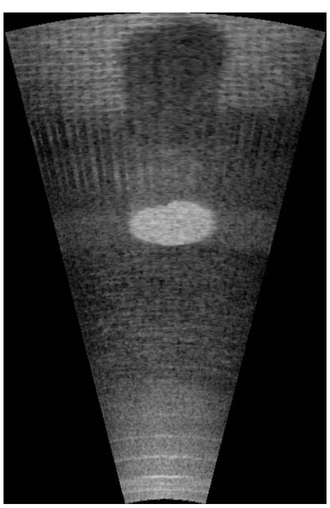

Fig.7 Concrete Cylinder Image

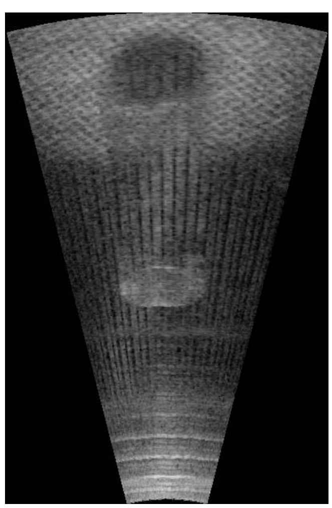

Fig.8 Metal Cylinder Image

interest, Fig.7 is an example of the situation displayed in Fig.14.

*2) Object surface and shape sensitive image*: The strength of the objects return *I* is dependent on the objects surface composition and shape. Figs.7 and 8 demonstrate the effect of using objects with various surface compositions on the acoustic image. Both objects, cylinders, are of the same dimensions, but manufactured out of different materials, concrete and metal. The top of the coarse concrete cylinder is displayed on the acoustic image as brighter than the top of the fine metal cylinder. This is due to the strong backscattering from the rough concrete surface.

Shape also has a large effect on the acoustic images. An example can be seen by comparing the box seen in Fig.2 and cone seen in Fig.10. Both objects are composed of the same material, concrete, however of vastly different shapes. The acoustic image of Fig.2 can be seen in Fig.4 while the acoustic image of Fig.10 can be seen in Fig.12. Notably the Fig.2's flat surface can be clearly identified in Fig.4, however the cone seen in Figs.10 is much less clearly recognizable.

These examples illustrate, that due to differences in surface composition and shape, different objects different recognition strategies.

*3) Changing features:* The acoustic reflection of the object is not stable as optical reflection and, as a result, the intensity of the acoustic returns are prone to change. As a result objects edges and areas are very noisy and difficult to detect. The images seen in Fig.15 demonstrate this effect. Both images where taken successively and are of the same object, a cone (seen in Fig12). A simple threshold was applied to the images and the images were binarized. As can be observed the edges of the object are show large amounts of noise.

*4)3D image*: The acoustic camera provides a 3D image and very hard to provide a 2D due to its displaying mechanism. For a human, 3D images are familiar and easier to understand than 2D images. However in machine vision, 3D image recognition is difficult and has not reached a practical level.

Due to the mentioned differences, the optical camera model and conventional recognition technique are difficult to apply to the acoustic camera image. As a result, to develop an effective image recognition algorithm, an acoustic camera model and recognition method had to be developed that considered the acoustic reflection characteristics of the acoustic camera.

## III. Proposed Method

In order to overcome the above difficulties and implement to AUV, we propose the camera model and the recognition method.

### *A. Modeling*

The image obtained by the acoustic camera is highly sensitive to the camera's position. The camera model allows us to predict the actual shape of the target object based on the image obtained for any given arbitrary position of the camera.

Fig.16 illustrates the acoustic camera's data collection and display method. The orientation of the camera beam can be determined by the two angles $\theta$ and $\phi$, defined in the

Fig.9 Concrete Tapped Cylinder
(15.5(D)x10(D)x5.8(H)cm)

Fig.10 Concrete Cone
(21(D)x25.5(H)cm)

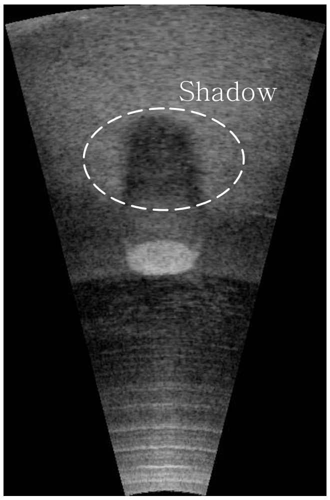


Fig.11 Concrete Tapped Cylinder Image

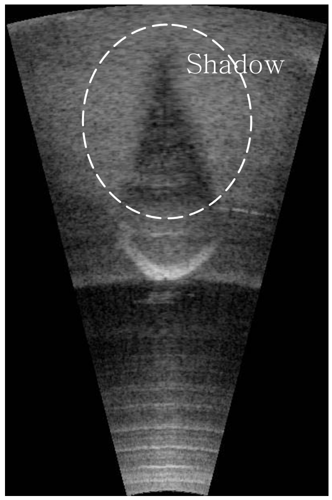


Fig.12 Concrete Cone Image

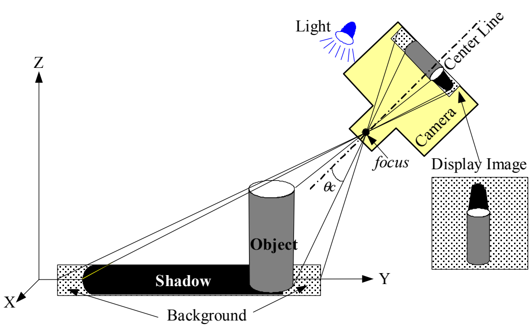


Fig.13 Optical Camera's Display Method

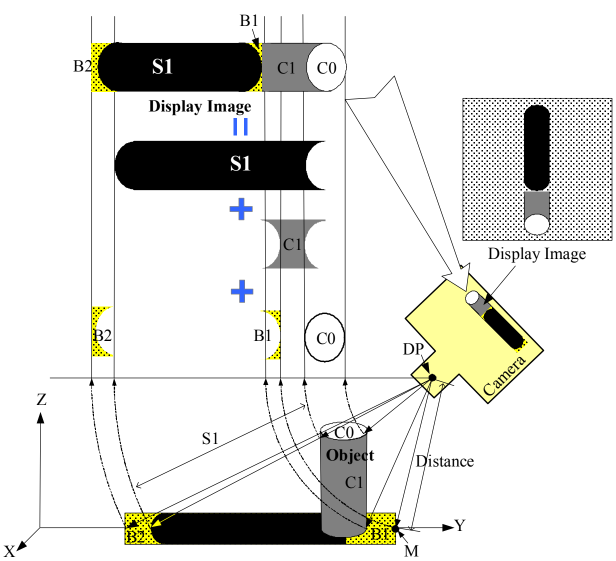


Fig.14 Acoustic Camera's Display Method

spherical coordinate system. As the camera tilts and pans (rotation on the XcZc and XcYc planes, respectively) the beams can be projected in varying directions and distance and intensity data collected.

Discrete tilting and panning angles with regular increments are used in the actual implementation. The tilting of the acoustic camera generates a vertical beam slices consisting of *N* number of tilted beams. Each beam returns a distance *D* and intensity *I*. The distances and intensities are then mapped to a single line in the camera image. By panning the vertical beam a series of lines can be generated that is used to construct a whole image. Assume, as shown in Fig.17, an object is located in the camera's visible range. Let a point *T*, on the object, set the origin of the image. Let the position and rotation of *T* be known. A beam *B* hits the object surface at the point *K*. The beam *B*'s pan and tilt angle from the center line are $\phi_b$ and $\theta_b$, respectively, and the length(range) of the *B* from point *C0* to point *K* is known. This describes the line *B*. Assume that the object's surface can be described the collection of triangular elements and the line *B* hits at the surface of element *S*.

The intersection of the *B* and *S* determines the point *K*. The distance *D* can be estimated using the distance between points *K* and *Co*. The intensity depends on object's surface composition and shape. The surface composition is, in turn, related with acoustic reflection characteristics *ma*, and the reflection angle $\alpha$ between the normal vector of the *S* and *B*. *I* can be estimated using the function $I_f(ma,\alpha)$ at *K*.

Using this method, each beam's distance *D* and intensity *I* can be estimated and used to construct the whole image frame as shown in Fig.16. This enables to predict an object image in the camera at a certain point *T*.

When an object image is taken, the camera position (Tx,Ty,Tz) and the rotation (Troll,Tpitch,Tyaw) can be estimated. Tx and Ty can be estimated using the position of an object in the image plane. Since all beam's $\phi_b$, $\theta_b$ and *D* is known, Tz, Troll and Tpitch can be estimated using triangulation. Tyaw is related with the camera's relative heading angle to a certain direction of an object. Due the large amounts of noise in the acoustic image, ellipse approximation [12] has been found to be the most efficient method of estimating the object's heading angle in the image. Using this camera model, the corresponding target image model can be used for the recognition.

*B. Recognition Cues*

Generally, the optical camera uses the highlight area of an object as a cue to recognize the object. Most often, the shadow area of the object is not used.

In the acoustic camera's case, shadow areas generated turn out to be the more general and reliable cues that can be used to recognize the object.

As shown in Fig.1 through Fig.12, highlight areas that are easy to recognize are only generated under the limited conditions, such as when the there is a flat and coarse surface. However, all objects which have heights generate a shadow

and generally most seabeds give good enough contrast to detect shadows.

Figs.15 Sucessively Captured Images of the Shadow of Fig.12 Cone

Table 1 Available Cues for Recognition

| Surface Status / Shape | Course | Fine |
|---|---|---|
| Polyhedron | Clear highlight area , Shadow | Dim highlight area, Shadow |
| Curvature | Shadow | Shadow |

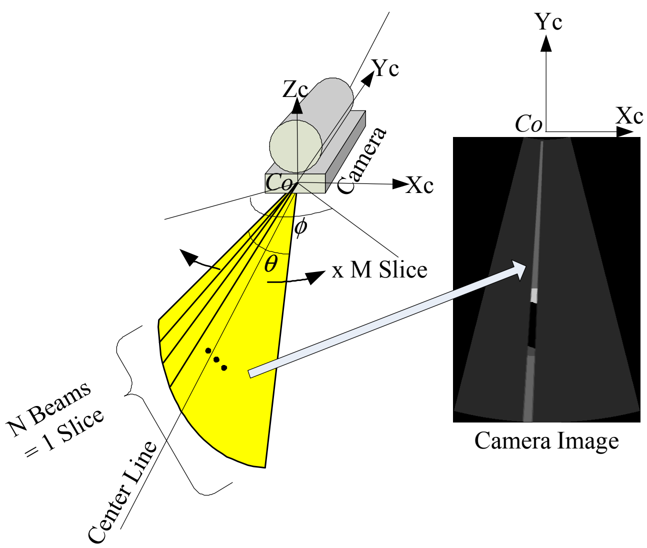


Fig.16 Acoustic Camera's Beams and Display

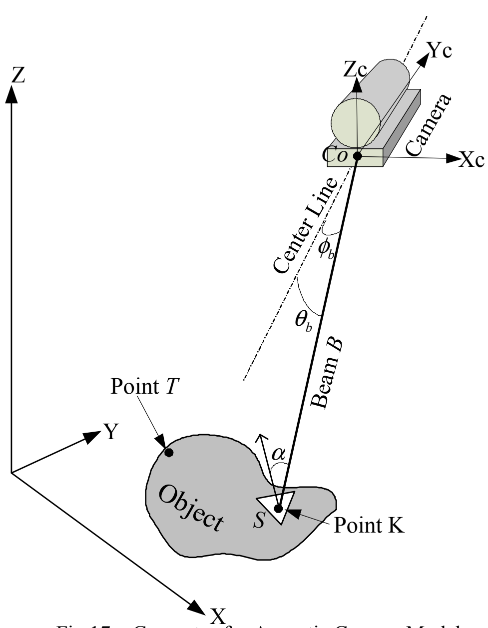


Fig.17 Geometry for Acoustic Camera Model

Table 1 groups the available cues of objects based on the experiment results. For example, the objects in Figs. 1 and 2 have coarse surfaces and are polyhedrons. As a result, based on Table 1, two cues that can be used in recognition, a clear highlight area and a shadow. Using this table, available cues for given types of objects can be selected for the recognition.

### *C. Recognition Method*

As a recognition method we propose to use the objects shadow for recognition. By comparing an objects shadow with a predicted shape recognition can be made. Acoustic shadows are less dependent on acoustic refection, and as a result more stable and reliable. The shadow is recognized using the correlation of the actual and simulated shadows' X and Y Axis projection [11]. Let correlation of X-axis projection between the model and the actual image be CorrXp and Y-axis be CorrYp and the length of the object at X and Y axis are Lx,Yx.

The object correlation value *Ct* can be expressed using the following equation:

$$Ct = K_1 \times CorrXp + K_2 \times CorrYp \tag{1}$$

$$K_1 = \frac{Lx}{Lx+Ly}, K_2 = \frac{Ly}{Lx+Ly} \tag{2}$$

The longer length of the shadow the longer projection values exist and the more information is provided. This method is robust to the edge and/or inner area noise (such as seen Figs.15) and efficient enough to realize the real-time processing with small computing power.

## IV. Simulation

As examples, representative polyhedron and cylindrical objects' shapes were simulated based on the proposed camera model. Two cube shape objects, a brick (Fig.1) and a cube (Fig.2) and three cylindrically shaped objects, a cylinder (Fig.5), a tapped cylinder (Fig.9), and a cone (Fig.10).

The camera position, *T* of the objects seen in Fig.17 are, (0 cm,70 cm,79 cm) for the brick see in Fig.1, (0 cm, 57 cm, 80 cm) for the cube seen in Fig.2, (0,57,80) for the cylinder seen in Fig.5, (0,75,105) for the tapped cylinder seen in Fig.9 and (0,53,80) for the cone seen in Fig.10.

The camera's roll and pitch angle was 0 degrees and pitch(down facing) angle was 45 degrees respectively. The DIDSON camera's $\theta$ and $\phi$ in Fig.16 were 14 and 28.8 degrees, respectively. The intensity of the object's surface return was predefined at a certain value and the inside of a surface area (curved or flat) was assumed to have the same intensity.

Figs.18 through 23 illustrate the simulated models. The cube shape objects' highlight and shadow area both displayed as the actual image.

## V. Experiment

In order to estimate the accuracy of the camera model and the proposed recognition, experiments were carried out. The

goal was to recognize 5 simulated models (brick, cube, cylinder, tapped cylinder, and cone) using the actual images.

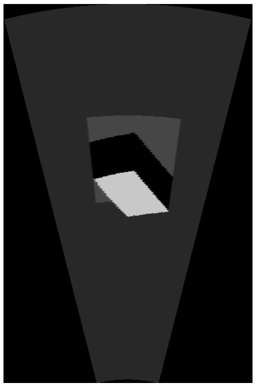

Fig.18 Simulated Brick Image of Fig.1

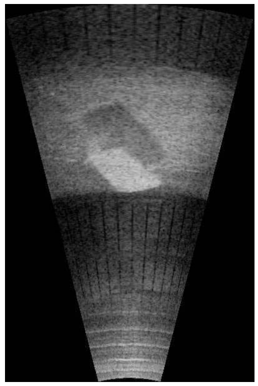

Fig.19 Actual Brick Image

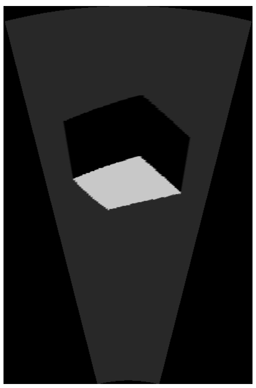

Fig.20 Simulated Concrete Cube Image at Fig.2

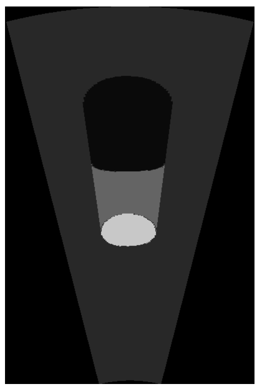

Fig.21 Simulated ConcreteCylinder Image at Fig.5

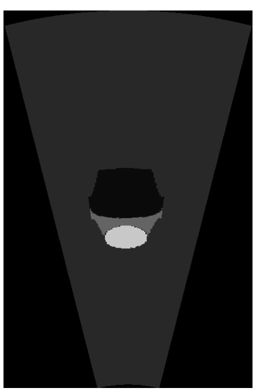

Fig.22 Simulated Concrete Tapped Cylinder Image of Fig.9

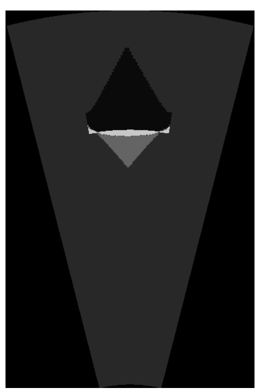

Fig.23 Simulated Concrete Cone Image of Fig10

The camera position $T$ and the rotation were the same as the simulation and the shadow area was used for recognition. As a recognition method, the proposed correlation method in section 3 was used.

Tables 2 and 3 illustrate the recognition results. The object's image was taken by the DIDSON with regular interval and the different actual images of the same object were used at each table to test the changing feature effect in the section2.

All objects were successfully recognized. The correlation result, $Ct$, was highest for when the simulated model was matched with the actual image of the object in question. The cone and the tapped cylinder showed quite similar values.

Table 2 Recognition Result1, Correlation Values

| Model / Actual | **Brick** | **Cube** | **Cylinder** | **Tapped Cylinder** | **Cone** |
|---|---|---|---|---|---|
| Brick | **0.937** | 0.871 | 0.814 | 0.811 | 0.786 |
| Block | 0.847 | **0.979** | 0.878 | 0.885 | 0.830 |
| Cylinder | 0.781 | 0.874 | **0.982** | 0.889 | 0.868 |
| Tapped Cylinder | 0.743 | 0.851 | 0.899 | **0.986** | 0.953 |
| Cone | 0.784 | 0.861 | 0.912 | 0.961 | **0.989** |

Table 3 Recognition Result2, Correlation Values

| Model / Actual | **Brick** | **Cube** | **Cylinder** | **Tapped Cylinder** | **Cone** |
|---|---|---|---|---|---|
| Brick | **0.979** | 0.886 | 0.822 | 0.833 | 0.796 |
| Block | 0.878 | **0.982** | 0.901 | 0.878 | 0.827 |
| Cylinder | 0.818 | 0.882 | **0.985** | 0.904 | 0.872 |
| Tapped Cylinder | 0.792 | 0.867 | 0.894 | **0.981** | 0.945 |
| Cone | 0.825 | 0.869 | 0.913 | 0.963 | **0.988** |

This was due to the similarities seen in their shadows. These results confirm the accuracy of the proposed camera model and recognition method.

When the object is not symmetric, the object's angle needs to be estimated and used with the correct model to generate a recognition method. The object's angle estimation experiment was carried out to evaluate its accuracy. The mentioned ellipse approximation method at the section 3 was used. As shown in Fig.24, an object, a brick was installed on the turn table and rotated from 0 to 360 degree. The brick's highlight area was used for the angle estimation. Every 30 degree, the brick's image was taken and its angle was estimated. The simulated brick's angle was estimated in 1 degree intervals. Fig.25 shows the result. The actual and simulated results showed the good agreement.

The estimated angles in Y axis were changed from -30 to 30 degree while the actual rotation in X axis changed from 0 to 360 degree. This phenomenon is due to the slant of brick with respect to the camera and can be compensated for using the geometry and knowing the slant angle. Around 270 degrees, a large error can be observed. At 90 and 270 degree, the brick's shape becomes very similar to a square when viewed at a slant. In this case, the ellipse approximation is difficult since the length of major and minor is similar.

This phenomenon can avoided by finding the ratio between the brick's X-axis projection length with respect to the Y-axis projection length to find these angles. As shown in Fig.26, the X/Y ratio approaches 1.0 near of 90 and 270 degrees for the actual and simulated cases.

## VI. Conclusions

We proposed an acoustic camera model and recognition method. The recognition method considered acoustic characteristics of the DIDSON camera. As a result, high accuracy and reliability of this recognition method achieved and proved experimentally.

Optical camera models enables prediction object shapes and shadows at certain points. This plays an important role object recognition.

In this paper we proposed a recognition method using object shadows for use with acoustic camera images. This recognition method allowed for reliable recognition with minimal computing power.

Using the camera model and study of the intensity function, which related with object shape and material, a complicated object can be simulated with high accuracy.

Side scan sonar's displaying mechanism is very similar to those generated by the DIDSON acoustic camera. The proposed camera model and shadow recognition algorithm can be applied to recognize side scan sonar images of object or seabed elevation maps.

## REFERENCES


[1]Son-Cheol Yu and T.Ura, " Experiment on a system of Multi-AUV Interlinked with a Smart Cable for Autonomous Inspection of Underwater Structures", International Journal of Offshore Structures and Polar Engineering, vol.14.no.4,2004, pp.274-283.

[2] F. Hover, J. Vaganay and M. Elkins,"Relative Survey Under Ship Hulls", Proc. of 14th International Symposium on Unmanned Untethered Submersible Technology CD-ROM, 2005.

[3] Hayato Kondo, et. Al.,"Fully Autonomous Observation of Breakwaters by an AUV Kamaishi Bay", Proc. of 14th International Symposium on Unmanned Untethered Submersible Technology CD-ROM, 2005.

[4] Japan Acoustic Committee, *Ultrasonics Handbook*, Maruzen Press, 1999. (in Japanese)

[5] Son-Cheol Yu,T.Ura and T. Sakamaki,"Recognition of Sacrificial Anodes using Parallel Lighting for Full Automatic Underwater Inspection by Autonomous Underwater Vehicles",Proc.of IEEE/MTS OCEANS, 2002, pp.291-298

[6] E. O. Belcher, W.L. J.Fox and W.H. Hanot;"Dual-Frequency Acoustic Camera: A Candidate for and Obstacle Avoidance, Gap-Filler, and Identification Sensor for Untethered Underwater Vehicle", Proc. of IEEE/MTS OCEANS, 2002, pp.2124-2128

[7] WWW.DIDSON.COM

[8 ] K. Kim and N. Neretti, "Non-Iterative Construction of Super-Resolution Image from an Acoustic Camera Video Sequence", Proc. of IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2005, pp.105-111.

[9] S. Negahdaripour, P. Firoozfam and P. Sabzmeydani, "On Processing and Registration of Forward-Scan Acoustic Video Imagery", Proc. of the 2nd Canadian Conference on Computer and Robot Vision, 2005, pp.452-459

[10] T. Kazuo and T. Hmasahiro, *Fundamental of Oceanography*, Sanseido Press, 1997, pp.155-157 (in Japanese)

[11] B.K.P. Hron., *Computer Vision*, MIT Press, 1986

[12] Robert M Haralick and Linda G.Shapiro, *Computer and Robot Vision vol I* , Addison-Wesley 1992, pp.642-648


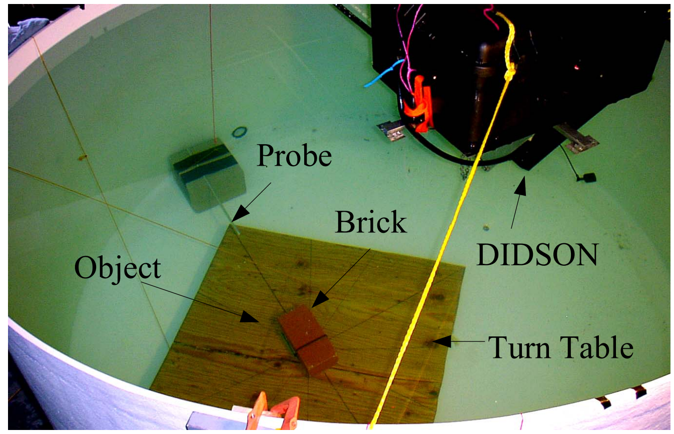


Fig.24 Object Rotation Test Experiment Setup

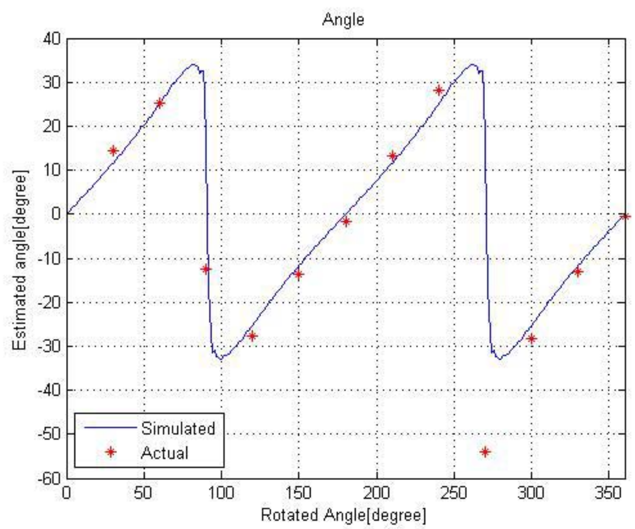


Fig.25 Rotation Angle Estimation Result

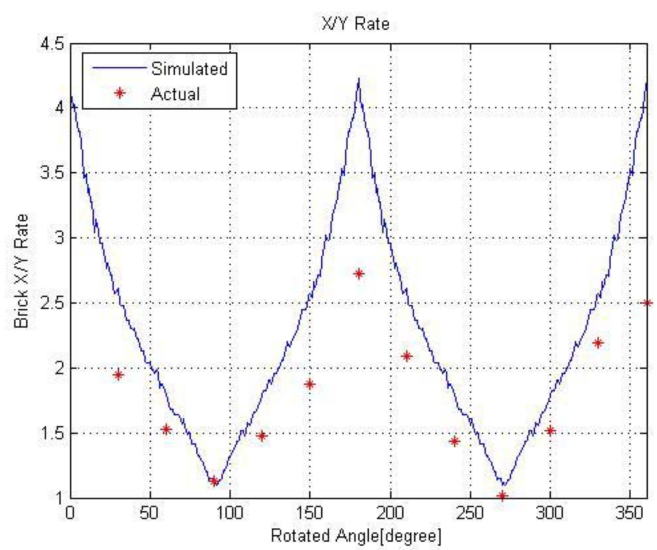


Fig.26 Brick's X /Y –Axis Projection Length Ratio